BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37351
doi: 10.3897/biss.3.37351

OPEN ACCESS

Conference Abstract

Applying Vocabularies to a Large, Historic Geoscience Collection

Adam Mansur‡, Leslie Hale‡

‡ Smithsonian Institution National Museum of Natural History, Washington, DC, United States of America

Corresponding author: Adam Mansur (mansura@si.edu)

Received: 15 Jun 2019 | Published: 26 Jun 2019

Citation: Mansur A, Hale L (2019) Applying Vocabularies to a Large, Historic Geoscience Collection. Biodiversity Information Science and Standards 3: e37351. https://doi.org/10.3897/biss.3.37351

Abstract

Geoscience collections have data needs that are broadly consistent with, but differ in important ways from, biology collections. Non-paleontology geoscience collections have received particularly little attention because they do not fit easily into the biodiversity framework that encompasses most natural history collections. Here we describe efforts to apply controlled vocabularies to the rock, mineral, and meteorite collections of the Department of Mineral Sciences in the Smithsonian Institution's National Museum of Natural History. Controlled vocabularies encourage consistency in usage of terms, connect related terms and concepts, and simplify discovery and re-use of data. Developing and implementing vocabularies for geoscience collections is a key step in better integrating these collections into the broader natural history landscape.

The Mineral Sciences collections contain approximately 450,000 specimens collected over 150 years. Specimen records therefore contain a large number of archaic terms and vary widely in the quality and amount of descriptive metadata available. We targeted four fields judged most likely to benefit from using a vocabulary: rock/mineral classification (using terms from the British Geological Survey Rock Classification Scheme, the International Mineralogicial Association's list of minerals, and other sources); rock microstructure (using terms from GeoScience Markup Language, or GeoSciML); stratigraphy (using MacroStrat); and collection locality (using terms from GeoNames and the Global Volcanism Program).

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

This process highlighted many of the challenges of grafting vocabularies onto existing data of varying age and quality, including:

- A lack of comprehensive, widely used vocabularies for geosciences
- Ambiguities in usage between the specimen database and the controlled vocabulary
- Widespread use of archaic terms that must be defined or replaced
- Difficulty shoehorning vocabularies into databases not designed to accommodate them
- Tension between the needs of data management and collection management

Even with these difficulties, implementing vocabularies has made it easier to retrieve and interpret specimen data, especially for queries of rock/mineral classification and locality information, and is a useful step in encouraging wider use of data from the collections.

Keywords

earth science, controlled vocabularies

Presenting author

Adam Mansur

Presented at

Biodiversity_Next 2019